DOCUMENT RESUME

ED 191 450 IR 008 682

AUTHOR Gawaldo, Rich: Schlueter, Betsy
TITLE Computerizing Internal Information in a Special Library.
PUB DATE 79
NOTE 12p.: Paper presented at the National Online Information Meeting (New York, NY, March 25-27, 1980).
AVAILABLE FROM Rich Gawaldo, Mobay Chemical Corp., Parkway West, Pittsburgh, PA 15205 (Free).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Programs: *Data Bases: *Information Retrieval: Information Storage: Information Systems: *Online Systems: *Search Strategies: *Special Libraries

ABSTRACT

Modeled on public databases such as Lockheed and SDC, the information system described in this report was utilized in the creation of in-house databases in the Mobay Chemical Corporation's special library in Pittsburgh. Using IBM's Storage and Information Retrieval System (STAIRS), Mobay created two databases: one to hold product information, the other to contain the titles of journal articles arriving in the library. The storage and retrieval capabilities of each database are detailed with regard to searchability, response time, Boolean logic capacity, updating, thesaurus control, and other considerations. (FM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED191450

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

COMPUTERIZING INTERNAL INFORMATION IN A SPECIAL LIBRARY

by

Rich Gawaldo
and
Betsy Schlueter

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Betsy Schlueter

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR008682

COMPUTERIZING INTERNAL INFORMATION IN A SPECIAL LIBRARY

by Rich Gawaldo and Betsy Schlueter

At Mobay Chemical Corporation's Pittsburgh headquarters our special library staff is much involved in the on-line searching capabilities which have so improved total information retrieval during the past few years. We began, in 1974, to search databases offerred by on-line vendors such as Lockheed and SDC. In the middle of 1979, we started to create our own in-house databases. I am here today to speak about our internal databases, what they can offer a Special Library, and our experiences with them at Mobay.

The software package that we are using is IBM's STAIRS system. STAIRS (Storage and Information Retrieval System) is a packaged set of programs which offers the user a variety of resources for database creation, search, and retrieval. The user queries are formatted in Boolean Logic and/or by means of comparison operators. These queries can be entered through a CRT-type terminal or keypunched and submitted on cards. STAIRS may be used in all areas where access to, and fast scanning of, large numbers of articles, reports, contracts, laws, general directives, or abstracts of publications are essential. Up to 15 databases can be concatenated and searched.

STAIRS operates on an IBM 370 computer or IBM equivalent via a control program. At Mobay, this program is IBM's CICS (Customer Information Control System), however, it can also run under IBM's IMS (Information Management System). STAIRS software and the control program cost about $2000. per month. Hardware costs vary with the installation. STAIRS can be accessed by any geographic

5.28

location defined to the system.

Data can be entered into STAIRS in two ways. One method is through ATMS (Advanced Text Management System), another IBM package. ATMS is a text processing system which has the primary function of creating and editing text but can also submit documents to STAIRS in a batch or on-line manner. Information can be entered into STAIRS in a batch manner via a user-written program, if you do not have ATMS.

The storage capacity of STAIRS is dependent upon the number of disk storage units associated with the system. At Mobay, disks are copied onto magnetic tape every night as part of the nightly backup procedures. Thus, documents can be restored in case of a system failure.

We currently have two databases up and available for on-line searching, although there is still some information to be entered into one of the databases.

The STAIRS software was originally considered by our special library staff and others as a way to handle a large amount of information that had been collected to comply with the Toxic Substances Control Act. The Act required that chemical producers and users report to the EPA (Environmental Protection Agency) the chemical composition of all the chemical products they manufacture or import. Other statistics such as production volume, production sites, descriptions of raw materials and chemical intermediates were also necessary. It was necessary to keep accurate records of all the data items so that we could identify specific products on the EPA's Toxic Substances Control Act Chemical

Substance Inventory and future supplements to that inventory.

The first database we created, which we now search via STAIRS, is our product information database. It was designed to hold the information collected under TSCA, and also to be used as an index to other internal information on our products. To date, we have a portion of the data collected for our Polyurethane Division up and searchable.

There were several reasons that a good system had to be developed for the storage and retrieval of this data: 1) the value of the data itself, 2) the large amount of data collected, 3) the number of different questions that the information in the database could answer if each data element is searchable, and 4) the number of different people that the database can answer questions for, e.g., customers who want to know whether or not a certain product has been registered with the EPA, employees who need to know the chemical composition of a certain tradenamed product, U.S. government agencies who want to know whether a certain product has been registered, and others who want to know the CAS registry number for a given product so that the public databases can be searched for information on that product.

The models for this information system were the public databases such as Lockheed and SDC which we access via our on-line computer terminals. The free-text searching capability of these databases was considered a key characteristic. STAIRS offers that capability.

Data in this database had to be current and accurate. The computer system had to be able to handle frequent changes in the data.

Our information was first gathered and verified on hardcopy forms. We designed questionnaires to obtain the initial product data from the many product managers in the three Pittsburgh Divisions. The questionnaires were sent out, filled out by the product managers,

and returned. Then we reviewed the information and copied it onto larger sheets.

We also reviewed import records and manufacturing processes to identify raw materials, intermediates, and by-products. We got the production and importation figures from other documents. And we recorded necessary data elements on large summary sheets with one sheet per product. We interviewed each of the product managers to be sure the information was accurate. Then CAS numbers were located for as many products as possible by searching the printed Chemical Abstracts and on-line dictionaries. Later, we filled out the appropriate EPA forms, and submitted them.

The manual system was adequate for the initial reporting. But soon it became clear that it could not handle the whole job. In many cases our specific products were considered mixtures under TSCA, and mixtures could not be registered. So instead of registering the product, we registered individual components of the product. If one of these components happened to be part of another product that was also a mixture, the production volume of these 2 identical components was totaled, and reported only once to the EPA. We had difficulty keeping track of this type of reporting with the manual system.

Also, after the EPA published its Inventory, we had to prepare new summary sheets to add the volume and page numbers of our chemicals in the published Inventory to our original data. These hand-written sheets soon became too small to hold all the data. Many became torn, and disorganized, and full of additional notations.

Over 900 different chemicals were reported company-wide, together with their raw materials and intermediates. The volume of the information that I had to check as different people asked questions was great. These dificiencies in the manual system promoted interest in computerized systems.

The characteristics of the information system we wanted to handle this data were: 1) completely searchable data elements, 2) immediate response time, 3) Boolean Logic capability, 4) easy updating, 5) thesaurus control, 6) minimum assistance from computer systems personnel once the databases were up and operating, 7) a small amount of system downtime, 8) good security because the database includes confidential company information, 9) several different possible output formats, 10) hardcopy print capability, 11) small physical space requirements, and 12) effective and simple processes for both inputting data and retrieving data.

There are many software packages on today's market that provide these system capabilities (e.g. Stanford University's SPIRES; BRS's, Lockheed's, and SDC's private files; THE SOURCE software; IN MAGIC; and others). The decision about which to use rests on costs and other applications you have for the software. In our case, costs were calculated for an anticipated database by estimating storage requirements and the number of records. Time constraints and hardware compatibility were also considered.

The fact that the computer system, STAIRS, was already available within the Corporation, and that the library could access it also and share costs, made it seem even more attractive. Also anticipating that more databases would be added to the internal databases now on-line for the library, STAIRS seemed most economical. There is very little added cost per database. Also, additional databases can be concatenated to the databases that are up. This means that if different databases contain identically formatted data elements, then those databases can be cross-searched.

For the product database, first an input form was created

on ATMS (Advanced Text Management System) by systems personnel. The input form was set up with a descriptive prompt (written in English) for each data element that will be searched on STAIRS. The inputting operator enters the information corresponding to the prompts and, after completing the form, stores it as a document. The stored document is then run through a series of commands which convert the prompts into paragraph classes which are to be interpreted by STAIRS to place the information into its proper place in the STAIRS document. Once the document is in this form it is passed to STAIRS. This conversion procedure was set up by the computer department and is accomplished by the execution of a single command.

Then the 26 paragraph classes or data elements that are possible with STAIRS were decided upon. They included product tradename, CAS number, chemical composition, production site, etc.

I decided to use the CRT data entry via ATMS for this product-related database to achieve accuracy, and also because the information was not already in machine-readable form. Several of the data elements were designated as formatted data elements which meant that they were limited to a certain number of characters. These data elements included CAS number, product code, etc. One advantage of having designated formatted fields is that changes can be made instantaneously to these fields while searching in STAIRS. This eliminates any inaccuracies in information immediately. You can also sort on formatted fields. Another advantage is that the formatted fields alone can be shown as one of the viewing options. Each of our documents, for the most part, were more than one screen, so this was an important feature. Then the variable fields can be viewed separately, or the total document can be viewed. There is no limit to the amount of information allowed in a given data element.

Terms can be combined for searching, using and, or, not, adjacent, same sentence, same paragraph logic. Truncation is possible and the various root terms are automatically shown when truncation is used, similar to what you see when you use Lockheed's expand command.

An alphabetically ordered printed list of terms in this database was automatically created to assist searchers with synonyms.

Synonym tables are also searchable with STAIRS using the SYN=ON command. For example, if you want the additional term "colour" searched every time you entered "color" you could set up a table to do this. To date, we have not done this with the chemical synonyms and abbreviations, because we felt the tables should be limited to one-term words, not chemical names which are multi-term words.

There are some other nice features in STAIRS: 1) a SORT command for sorting up to 600 documents, 2) a HIGHLIGHT=ON command which highlights the search term in the retrieved document, 3) a NEWPAGE=ON command which will have each separate document printed on a new page, 4) a DETAIL=ON command which displays the number of occurrences for every word in a query, 5) a DISPLAY mode which allows a display of the search history, 6) MAIL and PRINT commands which allow various hardcopy prints of retrieved documents at various locations, 7) the PURGE command which deletes all previously entered queries, 8) the RANK command which allows the searcher to weight the retrieved documents based on the frequency of occurrence of these words in the documents, and 9) the LIMITING command which limits the search term to a certain field. This latter feature is important in our product databases where we want to retrieve

the main document on a given tradename rather than all the other products that may contain that product as one of the components. We can do this by entering the search term and then .tradename. .

Both the manual system and the computer system were used in parallel while the data conversion was proceeding.

The full effectiveness of this database will be demonstrated in the future as the number of users grows. The system itself does have growth possibilities also. Additional terminals can be installed at various geographical locations. A variety of different types of information can be added to form additional databases. Searching across these different databases is possible using the SAVE, CHANGE, and EXEC commands.

Since the volume of data that can be entered into this system is virtually unlimited, additional data elements can be added to the database without any additional systems work.

Besides product data, our Special Library has had for several years several internal indexes in machine-readable form to create printed KWIC Indexes. Since STAIRS offers the capability of either CRT data entry through ATMS or data entry via IBM punched cards, this meant that indexes that were already in machine-readable form could be transfered easily into our STAIRS internal databases. Some of the KWIC Indexes have become very large over the past five years so storing them in the computer and searching them on-line was definitely viewed as an improvement. Storage space was saved and updating was easier.

The second database of the two that are now on-line is our "journal title database". It contains the titles of journal articles that arrive in the library each day. These titles are keypunched

when the journal arrives. The database is searchable by title words, and a journal coden which indicates the title of the journal by 2 alphabetic characters followed by six digits which indicate the month, day, and year of that issue.

This database was originally designed to accomodate those people who needed information on a given subject right away. Those patrons did not have time to wait for full references to be ordered after the citations were retrieved from the public databases. By entering the search terms into the internal database, references on a given subject could be retrieved for which the complete article is in our library. This helps people that must give rather spontaneous talks, or people that must meet customers to discuss a subject about which they would like additional information right away.

The database has other nice features. The journal coden is directly searchable. You can, if you want, scan all titles from a certain journal you receive to see the type of articles in the journal. This might even help you decide whether or not you want to continue subscribing to a given journal.

I have been tagging each title, also, to give it a subject category designation. Then after all the titles for a given update have been loaded into STAIRS, it is possible to search the subject tag and print out all the document titles and their journal codens on each subject. Then a list of those titles is circulated to the persons who are interested in each subject category. We have already seen instances where information from this alerting service has saved money.

The date of entry into the database of each group of journal titles becomes a part of each title and is a searchable data element.

Over 10,000 titles and their corresponding journal codens were loaded with the first update. That represents only a few months of journal titles. So you can see that this is rather a large database as far as the number of documents that have been entered, although the amount of data on each document is small. Some of the software vendors that offer storage and retrieval capabilities to outsiders charge per document stored. In this respect, STAIRS offered a lower cost to us because unless there is a hugh amount of data which would require an extra disk pack there are no additional storage fees. So far two separate divisions at Mobay have entered data into the computer via STAIRS and have not required an additional disk pack.

We hope to add more of our index and abstract files to the computer using STAIRS in the not-too-distant future as additional databases. Possible files to be added are our internal reports, analytical reports, retrospective search index, and book cataloging data.

Databases such as the two I have described in detail, and the others I have suggested as possibilities that contain a Corporation's internal information, together with the search capabilities that have already been proven for published information, should greatly enhance the information retrieval and resources of today's Corporations.